An approach to reduce SIFT computational cost (QSIFT)

***Abstract*—This paper describes the method to increase the speed of SIFT feature extraction by feature approximation instead of feature calculation in various layers. SIFT has been proven to be the most robust local rotation and illumination invariant feature descriptor. Additionally, it supports affine transformation. Being fully scale invariant is the most important advantage of this descriptor. The most major SIFT'sdrawback is time-consuming which prevents utilizing SIFTin real time applications. This research attempts to decrease computational cost withoutsacrificing performance. The recent researches in this area approved that direct feature computation is more expensive than extrapolation. Consequently, contribution of this research reduces processing time considerablywithout losing accuracy.**

Keywords-Interest points, Feature detector, Feature descriptor, Feature extraction, Feature matching, natural image statistics, real-time

INTRODUCTION(HEADING 1)

Feature detection and image matching are two essential steps in machine vision and robotics applications such as object recognition and matching [5], 3D scene reconstruction [6], motion tracking [7], image representation [8], image classification and retrieval [9], robot localization [10], texture classification [11] and biometrics systems [12]. A desirable feature detection methodmust be invariant to image transformations such as scale, illumination, rotation and affine transformations.

Scale Invariant Feature Transform (SIFT) is accurate and efficient in many applications speciallyobject detection and recognition [1]. SIFT major drawback in real-time application is computational complexity.

Speeded-Up Robust Features (SURF) includes detection, description, and matching steps [4]. However, SURF is faster than SIFT but its accuracy is less.

Binary Robust Independent Elementary Features (BRIEF) is another alternative for SIFT.Although theBRIEFcomputational costsless than SIFT but itsmatching rate is not comparable to SIFT in some applications [15].Another efficient alternative to SIFT or SURF is Oriented FAST and Rotated BRIEF (ORB) [16].SIFT is the most robust descriptor considering advantages and disadvantages of all mentioned descriptor in this section.

The rest of this paper is organized as follows: 2 describes related work,3 includes proposed methodology in this research,4 contains the comparison between the current used dataset in this area and developed datasets by this research and 5 represents experimental evaluation and finally further development and conclusions is available in 6.

RELATED WORK

Feature channel scaling

Feature channel scaling is a method that increases calculation speed feature extraction in down or up sample based on original image without losing a considerable degree of accuracy [2].

This method has studied the image behavior and multi scale's features. And observed since each single image inside a specific dataset includes small patches thus its behavior is similar to the dataset behavior. As it indicates in following formula [2].

$$f_{\Omega}(I_s) \equiv \frac{1}{h_s w_s k} \sum_{ijk} C_s(i,j,k) \quad \text{where } C_s = \Omega(I_s) \quad (1)$$

Assuming the Ω is any low-level shift invariant function and I is input image then C represents the new channel of image as: $C = \Omega(I)$

The several local and global features can be described by formula (1)[3]. If Is denotes input image (I) at scale s then:

$h_s \times w_s = s(h \times w)$

when $h \times w$ denotes dimension of I and $h_s \times w_s$ is dimension of Is. $\forall s > 1$, Is the higher resolution version of I, while $\forall s < 1$, Is obtained by interpolation of I.

In fact $f_{\Omega}(I_s)$ represented the global mean of Cs.

The relation between input image and its scales can be described by (2).

$f_{\Omega}(I_{s1}) / f_{\Omega}(I_{s2}) = (S_1 / S_2)^{-\lambda_{\Omega}} + \varepsilon \qquad (2)$

If ε denotes the error then the effort has been made to train $\lambda_{\square}$which lead $E[\varepsilon] \approx 0$. It is valuable to note that each channel type Ω has own λ_{Ω}.

All parameters in (2) are available except λ_{Ω}that has been learnt in learning phase then uses in (3).

Assuming s_1=s , s_2=1, by rearranging (2) ,(3) is obtained:

$f_{\Omega}(I_s) \approx f_{\Omega}(I) s^{-\lambda_{\Omega}}$

(3)

Beside for any corresponding windows w_s in I_s and w

in I, the following

g formula can be utilized:

$$f_{\Omega}\left(I_s^{w_s}\right)\approx f_{\Omega}\left(I^{w}\right).s^{-\lambda_{\Omega}}$$

$$C_s\approx R\left(C,s\right)s^{-\lambda_{\Omega}} \quad (4)$$

Considering λ_{Ω} is learnt in train phase then it can be used in test phase and approximated$C_s$according (4) with each scale. It allows general, simple and accurate fast feature construction.

SIFT

The main steps for SIFT descriptor extraction are as the following:

Scale-space extrema detection: it includes searching all scales and image locations in order to efficiently implementing it utilizing Difference-of-Gaussian (DOG) function to determine potential invariant interest scale and orientation.

*2) Keypoint localization:*location and scale can be determined by a detailed model of each candidate location, which their Keypoints are chosen based on their stability of measurements.

3)*Orientation assignment*: each keypointlocation based on local image gradient directions can be assigned by one or more orientations. An orientation, scale, and location related image data is applied to all performance of feature operations.

4)Keypoint descriptor: The region around each keypoint is chosen for measuring local image gradient, which has been transformed to comprehensive levels of representations for illumination and local shape distortion modification. *[1]*

METHODOLOGY

The first step of SIFT for identify potential interest points, applies Gaussian function to each scale in every octave, then these results lead to obtain DOG. An individual relation between the result of Gaussian function in first octave and othershas been observed by this research based on 2.A and modeling.

Dataset generation

To provide different states of an image under various scale, illumination and rotation a sensitized dataset has been generated. This dataset supports toapprove stated claim in II.A. A synthetic dataset includes 2000 images has been augmented with camera parameters.(K,R,T)

Lambda calculation

In order to obtain the specific λ for Gaussian channel (2) was changed to (5)and λ representsas $\lambda_{Gaussian}$.

$$f_g\left(I_s\right)/f_g\left(I\right)=s^{-\lambda_g}+\varepsilon$$

$$\log\left(\mathrm{f_g(I_s)}\right)-\log\left(\mathrm{f_g(I)}\right)=-\lambda_g\,.\,\log s+\log\varepsilon$$

$$\lambda_g=\log_s \mathrm{f_g(I)}-\log_s \mathrm{f_g(I_s)}+\log_s\varepsilon \quad (5)$$

Using generated images in III.A. andaims$E\left[\log_s\varepsilon\right]\approx 0$, $\lambda_{Gaussian}$was trained to 0.18.

$\lambda_{Gaussian}$timing process depends on size and the number of images in dataset. Besidesmachine software and hardware configuration affects performance .Since training λ is offline process therefore can be ignored in online applications.
Machine software and hardware configuration:
As stated before, trained λ is fix for eachchannel.

Utizing lambda to….

Using (4), calculated Gaussian in first octave and $\lambda_{Gaussian}$, the result of Gaussian in other octaves in each scale will be estimated. Equation (4) will change to (6).

$$C_s\approx R\left(C,s\right)s^{-\lambda_{Gaussian}} \quad (6)$$

Fig1. Demonstrations (6) visually.As shown in Fig.1 standard pipeline do 2 steps for calculate Gaussian in intended scales. First, change input image to target scale and has been shown by I_s. $s=\left\{\frac{1}{2},\frac{1}{4},\frac{1}{8},\ldots\right\}$. Second,apply Gaussian function in I_s.
Whiles, QSIFT pipeline apply Gaussian function in input image and obtain$C=G\left(I\right)$ just once. Then using C and $\lambda_{Gaussian}$result of Gaussian in each scalecan beobtain directly. This pipeline has 1 step and do not need to change scale of input image.

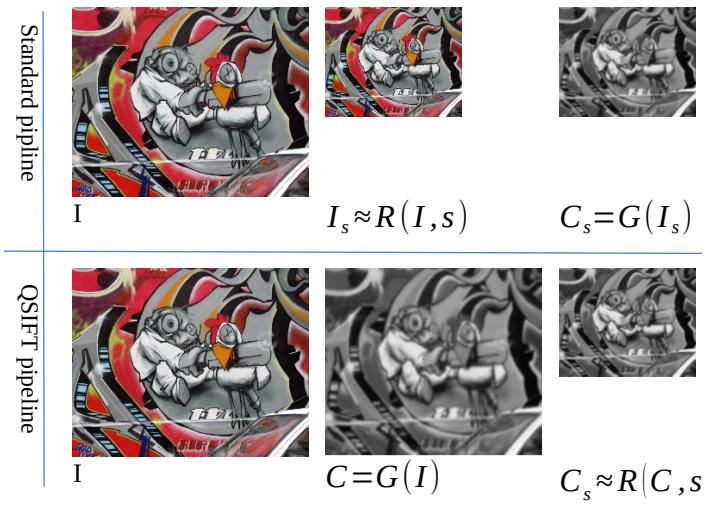


Figure 1. Top: the standard pipeline. We have to rescale image and then compute new channel. Bottom: we need compute new channel of input image for the first time and then approximate $C_s$in every scales.

Figure 2 illustrate standard pipeline. If standard pipeline includes 5 octaves with 6 scales in each octave, Gaussian must be applied 30 times per image.

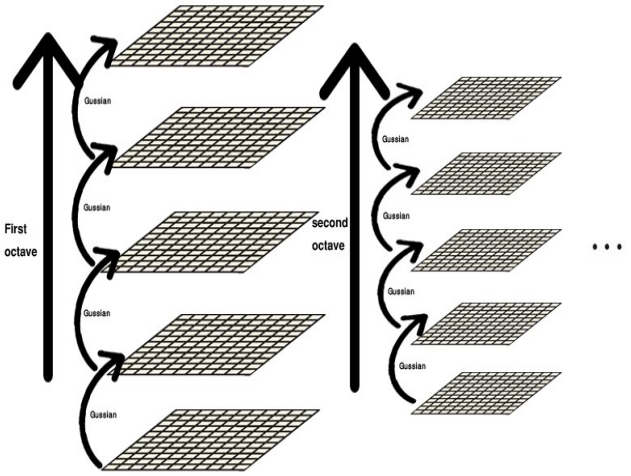


Figure 2. Standard pipeline

Figure 3 shows the proposed pipeline by this research that reduces the number of applying Gaussian function to 5.
The results of proposed method by this research establishes that time cost is significantly reduced if the first octave Gaussian supports to approximate others, avoiding computing Gaussian over all octaves.Since Gaussian is shift invariant function this hypothesis is expressed based on 2.A

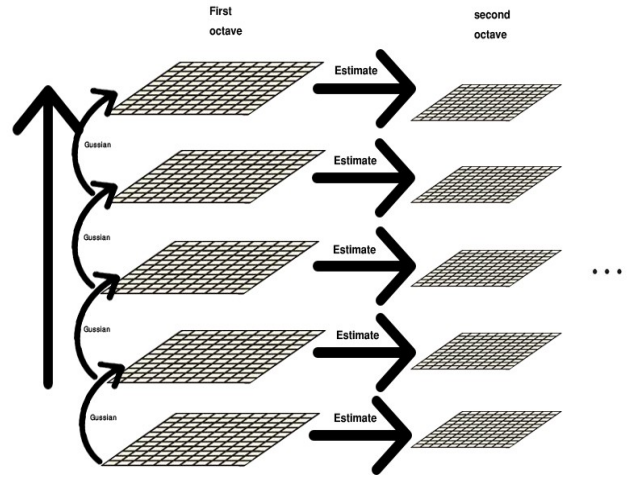


Figure 3. Proposed pipeline

Fig. 4. Left: original image. Right: simulate the image with camera with its deformation

EVALUATIONS

Generally to evaluate descriptors functionality several ordinary datasets are used such as ……..
For getting the better results and providing more various versions of an original image during this study a dataset was developed considering some parameters.
……..

10 images for each experimentation

TABLE I.

Table Head	Table Column Head			
	QSIFT rate	***SIFT rate***	***QSIFT matching count***	***SIFT matching count***
Scale				
Rotation				
Illumination				
Scale & rotation				
Scale & Illumination				
All parameters				
average				

a. Results of matching the image with its scaled image (based on time)

TABLE II.

Table Head	Table Column Head			
	QSIFT rate	***SIFT rate***	***QSIFT matching count***	***SIFT matching count***
Scale				
Rotation				
Illumination				
Scale & rotation				
Scale & Illumination				
All parameters				
average				

b. Results of matching the image with its scaled, rotation and illumination image (based on accuracy)

CONCLUSION

In this paper, we compared two different image matching techniques for different kinds of transformations and deformations such as scaling, rotation and illumination. For this purpose, we applied different types of transformations on original images and displayed the matching evaluation parameters such as the number of matching points in images, the matching rate, and the execution time required for each algorithm.

(1) References

[1].David G Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol.50, No. 2, 2004, pp.91-110,2004.

[2]. “Fast Feature Pyramids for Object Detection”

[3]. “Integral Channel Features”

[4]. “Speeded-Up Robust Features (SURF)”

[5]. Andreopoulos, A., Tsotsos, J. “50 years of object recognition: directions forward.” Comput. Vis. Image Underst. 117(8), 827–891,2013.

[6]. Moreels, P., Perona, P. “Evaluation of features detectors and descriptors based on 3D objects”,Int. J. Comput. Vis. 73(3), 263–284,2007.

[7]. Takacs,G., Chandrasekhar,V., Tsai, S., Chen,D., Grzeszczuk, R.,Girod, B. “Rotation-invariant fast features for large-scale recognition and real-time tracking”. Sign. Process. Image Commun.28(4), 334–344 ,2013.

[8]. Yap, T., Jiang, X., Kot, A.C.,” Two-dimensional polar harmonic transforms for invariant image representation”. IEEE Trans. Pattern Anal. Mach. Intell. 32(7), 1259–1270,2010.

[9]. Liu, S., Bai, X., “Discriminative features for image classification and retrieval”. Pattern Recogn.Lett. 33(6), 744–751,2012.

[10]. Murillo, A., Guerrero, J., Sagues, C., “SURF features for efficient robot localization with omnidirectional images”. International Conference on Robotics and Automation, pp. 3901–3907.Rome, Italy, 10–14 Apr, 2007.

[11]. Lazebnik, S., Schmid, C., Ponce, J., “A sparse texture representation using local affine regions”. IEEE Trans. Pattern Anal. Mach. Intell. 27(8), 1265–1278 ,2005.

[12]. Farajzadeh,N., Faez, K., Pan,G.,”Study on the performance of moments as invariant descriptors for practical face recognition systems.”, IET Comput. Vis. 4(4), 272–285 .2010.

[13]. Thao Nguyen, Eun-Ae Park, Jiho Han, Dong-Chul Park, and Soo-Young Min, “Object Detection Using Scale Invariant Feature,” Advances in Intelligent Systems and Computing, vol 238. Springer, Cham, pp 65-72,2014

[14].

[15]. Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua,” BRIEF: Binary Robust Independent Elementary Features”, Computer Vision – ECCV 2010, vol 6314, pp 778-792

[16]. Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski, ”ORB: an efficient alternative to SIFT or SURF”, Computer Vision (ICCV), 2011.

[1]

[2]

[2]

[3]

[1]

[4] ***Abstract*—The investigation surrounding recent Stockholm and New York terrorist attack enforced this research to emphasize on anomaly detection. This paper describes a main part of ongoing study through anomaly detection and localization which aims to improve offline/online accuracy. The sparsity constraint used in most recent anomaly detection researches was replaced with Locality-constrained Linear Coding. Locality-constrained Linear Coding reconstruction cost criterion is designed to detect anomalies that occur in video locally. Implementation this method and obtained experimental results approved considerable improvement regarding localization beside computing complexity reduction in dictionary learning.**

[5]

II. Introduction

[6] In recent years using Closed-circuit television video (CCTV) gets pervasive to prevent insecurity in crowded places. Traditionally uninterrupted monitoring required for scenes supervision, in the result operators face on some problems such as tiredness and carelessness, in critical conditions stated issues reinforce disaster occurrence probability. In addition, investigating large volume of daily generated videos is frustrating job [16]. Considering mentioned disadvantages of traditional supervision it has been noted implementation of automatic supervisions system is an extensive demand [2]. Computer vision techniques automatically analyze the stream videos to make alert when anomaly occurs [2]. The main target of automatic image supervision system is anomaly detection [16]. Anomaly behavior detection firmly depends on previous knowledge and human interfere in learning process [2]. There are three categories in learning process, for anomaly detection, 1) supervised [6-9], 2) semi supervised [1,2], 3) unsupervised [12-15]. Due to lack of sufficient training samples for anomaly detection in most cases unsupervised detection is performed. Generally, in this stage the system is trained with normal training samples then any incompatible models with train samples can be detected as anomaly [16].

[7] The rest of this paper represents the following sections:

Section II: related work, Section III proposed method, Section IV: Result and evaluation and finally Section V: Conclusion

III. Related works

[8] There are two methods for features extraction :1) Trajectory-based methods 2) Statistic learning- based methods.

A. trajectory based methods:

[9] These methods analyze trajectory based on clutter blobs tracking in consecutive frames [17-21], and makes a normal motion model then any derivation from trained model is translated as anomaly. However, these methods are accurate only for uncrowded scenes.

[10] The main drawback of these methods is that only anomalies spatial deviations have been noted and if an object moves normally, it has been ignored in anomaly classification without considering its appearance. In addition, these methods are failed when it needs to model crowded and complicated scene [6].

B. Statistic learning- based methods:

[11] To address these stated issues in section 2.A, researchers proposed Statistic learning-based methods. In these methods apply low-level features extracted from the pieces of the frame or spatio-temporal video volumes.

[12] These features include, optical flow [25,27], histogram of spatio-temporal gradients [16,17], etc. After feature extraction, statistical models such as Hidden Morkov Model (HMM), Bag of Word (BOW), Gaussian Mixture Model (GMM), etc. [1]are applied to detect abnormalities.

[13] [32] in sparse reconstructions methods to detect anomaly, normal bases are obtained and sparse reconstruction cost used to determine normality of input behaviors. [33] li et al proposed a method to detect locally and globally anomaly based on analyzing contextual information inside a video cube. They apply HOG3D descriptors to characterize motion patterns in these local volumes. codebook is generated to demonstrate globally atomic activity pattern and in order to describe locally salient behavior pattern inside an individual video cube, dictionary is constructed. eventually SRC criteria is used to detect anomalies.

IV. Proposed method

[14] The proposed algorithm in this research contains three modules:

[15] 1) low-level discovering (LLD) to make activity pattern codebook. 2) high-level discovering (HLD) to dictionary modeling of salient behavior. 3) Recognition video input data normality.

[16] This section briefly explains these three modules:

A. Atomic Activity Pattern Representation

[17] In order to anomaly detection first the video is described with features. This paper to describe a video uses spatio-temporal video volume[5]. Initially interest point in a video are recognized and then each point is considered as a kernel of a cube area with specific dimension which is called stvv. Then each stvv is represented by histogram of gradient (HOG) and histogram of optical flow (HOF) descriptors. To provide final feature vector HOG and HOF descriptors are concatenated. finally, each video can be represented as a collection of feature vectors, each feature vector is corresponded to an interest point. For

example, $X_k = \left(x_1 . \cdots . x_{n_k} \right)$, when n_k= the number of recognize interest point in a video. To represent an activity pattern with a collection of feature vectors a codebook is made for each video.

[18] This research is used fuzzy c-means clustering to make activity codebook in other to extract activity pattern frequently without considering locality.

B. Salient Behavior Patterns Representation

[19] Activity patterns are severely depending on their location. A normal activity pattern in a specific area maybe translated to anomaly pattern in another environment. For example, walking people in a gardening area of a park can be considered as an abnormal behavior, however walking is a normal behavior in pedestrian area. Therefor to salient activity discovery a local spatio-temporal context is proposed by this research. To implement this theory a video clip is divided to many video cubes with time and space overlapping. Then each video cube is converted to eight blocks. Stvvs are located inside block with collection of membership degrees, $U_j = \left(u_{ji} \right)_{i=1}^{N_c}$.All related membership degrees to a specific cluster are accumulated, thus the histogram vector for a block is obtained. All obtained histogram vectors are combined to provide final feature vector. The learned composition pattern can be represented as salient behavior pattern in local area however non similar composition representation vector can be considered as anomaly, due to restriction of trained samples and being high dimensional. Fitting a probabilistic model is challenging, eventually sparse representation is a suitable representation for high dimensional vectors. Training samples data set leads to extract a normal primitive set which is satisfy in $\phi = R^{m \times D}$.

C. Dictionary learning

[20] A trained sample can be represented as a linear sparse combination of primitive set. Nannan Li etall. [] proposed utilizing sparse coding with objective function to make optimum dictionary as it is shown in the following formula:

[21]

[22] $$\min_{D \in R^{d \times k}, \alpha \in R^{k \times n}} \frac{1}{n} \sum_{x_i \epsilon X} \left(\frac{1}{2} \left\| x_i - D\alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1 \right) (1)$$

[23]

[24] In this formula the first function is called loss and second one is called regularization.

[25] The loss function attempts to decrees error coding and regularization function guaranties the necessary sparsity of the x vector. To reduce value of formula (1) based on D,α frequently fixed one and minimizes another one until to make algorithm to be converge or minimize epsilon as epsilon<1. This research employed formula (1) to create training dictionary.To solve coding issue an optimization equation with L1-norm constraint must be solve however this equation solution leads to increase computing complexity. Although, L1-norm guaranties sparsity for a single sample, during reconstruction this sample just only uses some dictionary words that does not play any rolls in reconstruction similar samples. The lake of this ability to describe all samples causes to destroy sparsity. To address this issue LLC is proposed by this study. This method supposes that locality is more important than sparsity. The locality supports sparsity but it is not necessarily true for sparsity. The locally property of LLC method decreases reconstruction errors and increase ability to describe. To better understanding this property figure 1 illustrates locality in LLC method and lack of locality in SC. SC method to reconstruct two similar samples uses different words in a dictionary which is shown with yellow circle in figure 1. This attribute of SC method ruins the correlation between reconstruction coefficients, makes different reconstruction coefficients vector and thereby reduces the ability to describe two samples. In contrast, LLC method due to locality property for two similar sample generates two similar reconstruction coefficient vector and supports accuracy of locally anomaly detection. According to stated LLC functionality, this research eliminates sparsity constraint and replaces it with LLC. As it is shown in the followingformula:

[26]

[27] $$\min_{A} \sum_{x_i \epsilon X} \left\| x_i - D\alpha_i \right\|^2 + \lambda \left\| d_i \Theta \alpha_i \right\|^2 (2)$$

[28]

[29] Θ represent multiply vector elements, $d_i \in R^{|d|}$ indicates the Xi similarity with dictionary words.

[30] $$d_i = \exp\left(\frac{dist\left(x_i . D \right)}{\sigma} \right) \qquad (3)$$

[31] In this formula dist(x_i, D) means Euclideandistance, σ denotes normalizing coefficient.

[32]

[33]

[34]

[35]

[36]

[37]

[38]

[39] SC LCC

[40]

[41] Fig. 1. A comparison between SC and LLC, using words dictionary to describe xi.

[42]

[43] It is worth noting that vector α_i in equation 2 is not sparse as L0, and only a few elements in α_i have considerable values. Thereby for reconstruction rather than using all dictionary words, only low value reconstruction coefficients are converted to zero by threshold value. This paper for reconstruction is used k-nearest neighbors (kNN) to find nearest dictionary word to X_i vector in order to improve computing speed in equation 2. The reconstruction coefficients α^* are calculated as follows:

[44] $$\alpha^{¿} = arg\,min \frac{1}{2}\left\| x_i - D\alpha_i \right\|_2^2 + \lambda \left\| d_i \Theta \alpha_i \right\|^2 \qquad (4)$$

[45] α^* Calculation leads to obtain Locality-constrained Linear Coding reconstruction cost criterion to detect anomalies that occur in video locally.

[46] $C_x ¿ \left\| x_i - D\alpha_i \right\|^2 + \lambda \left\| d_i \Theta \alpha_i \right\|^2$. There is an inverse relationship between cost a normality.

[47] Assuming C_x has a great value ➔ a linear composition of dictionary words cannot reconstruct the input vector. In the result input vector with high probability is translated to be anomaly. Consequently, this algorithm can define a threshold value for cx which is called C_{th}. Eventually if $C_x > C_{th}$ then the input vector is represented anomaly.

V. MATH

[48] ion have been defined before the equation appears or immediately following. Italicize symbols (*T* might refer to temperature, but T is the unit tesla). Refer to "(1)," not "Eq. (1)" or "equation (1)," except at the beginning of a sentence: "Equation (1) is ... ."

VI. CONCLUSION

[49] The research undertaken applied a new approach to anomaly detection which applicable for CCTV in crowdedscenes. The proposed method detects local anomalies analyzing contextual information inside an individual video clip. Implementation this method on UCSD anomaly detection dataset and The experimental resultsthis confirmed improvement in efficiency and accuracy. further development will conduct to extend the offline in this research to real time system.

REFERENCES

[50] Basic format for books:

[1] D.Zhang,D.Gatica-Perez,S.Bengio,I.McCowan,Semi-supervisedadapted hmms for unusual event detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2005,pp.611–618.

[2] R.Sillito,R.Fisher, Semi-supervised learning for anomalous trajectory detection, in: Proceedings of British Machine Vision Conference(BMVC),2008, pp. 1035–1044.

[3] . AT. Xiang, S. Gong, Video behavior profiling for anomaly detection, IEEE Trans. Pattern Anal. Mach. Intell.30(2008)893–908.

[4] uthor, "Title of chapter in the book," in *Title of His Published Book, x*th ed. City of Publisher, Country if not

[5] USA: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx*.

[51] *Examples:*

[6] G.O.Young,"Syntheticstructureofindustrial plastics,"in *Plastics*, 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill,1964,pp.15–64.

[7] W.-K.Chen,*LinearNetworksandSystems*.Belmont, CA:Wadsworth, 1993, pp. 123–135.

[52]

[8] J. K. Author, "Name of paper," *Abbrev. Title of Periodical*, vol. *x,* no. *x,* pp. *xxx-xxx,* Abbrev. Month, year.

[53] Examples:

[9] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959.

[10] E. P. Wigner, "Theory of traveling-wave optical laser,"*Phys. Rev*., vol. 134, pp. A635–A646, Dec. 1965.

[11] E. H. Miller, "A note on reflector arrays," *IEEE Trans.Antennas Propagat*., to be published.

[54]

[55]

[56]

[57]

[58] A Method to Address Occluded Face Recognition in Authentication and verification

[59]

[60]

[61] Abstract—This paperinvestigates issues regarding multi factor authentication and verification based on face recognition under uncontrolled conditions such as illumination, expression, positions and occlusion and specifically studies surrounding occluded image In this research all affected facial images are supposed to be occluded, then occlusion parts of image are extracted and finally , occlusion mask added to normal frontal training face images. The Euclidean distance between synthetic occlusion image made by research compared to original occluded input image. The results indicates higher accuracy compare to current methods such as nn4 deep learning for face recognition.

[62] ***Keywords—occlusion; face authenticatin; face identification; face verification***

INTRODUCTION

[63] Generally themulti factor authentication system works based on the candidate identification through logging in with registered username and password (something the user knows) and recognizing him/her utilizing a biometric factor such as face(something the user is). In The multi factor authentication system, user must present a proof of presence to avoid cheating and fraud. This system increases the layer of security to electronic authentication

[64] Bio metric factor in this project is face and main drawback of using face biometric for identifications that it can never be hundred percent accurate. Two statistics are used to measure system accuracy: ref

[65] False Non Match Rate (FNMR) that means how often a biometric is not matched to the template when it should be

[66] False Match Rate (FMR) that means how often a false biometric is matched when it shouldn't be.

[67] Since making restriction for user to provide more than one a frontal normal facial image during authentication reduce system functionality, then the main target of this research is designing an accurate face recognition system under uncontrolled conditions such as illumination, occlusion, position and expression

[68] Multi factor Authentication includes two steps:

- Confirmation of the logged on informationcomparing registered information.
- Identification means the image of present user who is already logged in to the system is matched with the image of the claimed user.

[69] In addition, multi factor authentication module based on face recognition includes one pre authentication module (registration) and one post authentication module (verification) as the following:

- Registration: In biometric systems each user enrols by creating an account and providing some personal information and a photo which is used to make a template of that biometric. This image must be current, valid, and authentic. In most cases is making obligation for user to submit more than one normal frontal facial photo during registration or enrolment not practical and ethics, then designing a face recognition system with a small size dataset is required. In small size face recognition system dataset includes just one sample of.each class.
- Verification:Users can enter to a remote online secure systemas soon asbe identified through their username and password (E-authentication) The access to the system will be continued as long as successful verification results obtained by face recognition indicates that the logged in user is current, enrolled and constant during the access.

[70] The rest of this paper is organized to represent as the following:

[71] Section II: related work

[72] Section III:Methodology

[73] Section III: Experimental results and evaluation

[74] Section IV: Conclusion and further developments

RELATED WORK

[75] The most problematic challenges in face recognition is occlusion. SparseRepresentation based Classification (SRC) methods claims for high accurate and robust face recognition under occlusion. REF NO 12345

[76] **SRC : Robust face recognition via sparse representation**

[77] The following formulae are the basis of all mentioned methods:

[78] Considering $m \times n$is a frontal image then $d_{i,1} \epsilon R^{mn}$that represent jth samples from ith class.

[79] $D_i = ¿]k = number\ of\ samples \in each\ class$

[80] $D_T = [D_1, D_2, \ldots, D_l]$

$l = the\ number\ of\ classes$

[81]

[82] $y_{input} = D \times \alpha + \varepsilon,$

[83] Non occluded testing sample = y_{input}

[84] error $\varepsilon \cong 0$

[85]

[86] For occluded testing sample this method proposed offline making occlusion dictionary (*O)* and used it following equation.

[87]

[88] Occluded testing sample= $y_{input} = D \times \alpha + O \times \beta + \varepsilon$

[89] and error $\varepsilon \cong 0$

[90]

[91] Considering D_T is dictionary of trained samples and all SRC family method make it seamlessly without learning by reshaping and concatenating all samples in train dataset however they proposed their own approach to make *O*. According to published result by these researches SOC is the most accurate one.as it shown in fig 15.REF for fig

[92]

[93] Since verification in nature is a real time (online) process. Thus the system does not have any knowledge of type of occlusion in the result making occlusion dictionary is not feasible. Therefore this research supposes any input test sample is occluded and tries to extract occlusion mask from it. In the next step occlusion mask will be applied to all members of all classes in train dataset to generate synthetic occluded face. Finally the distance between HOG features of input test sample and each synthetic occluded trained data is calculated, the shortest distance represents the class of input sample.

[94]

METHODOLOGY

[95] Generally, any face recognition system regardless being offline or online contains the following submodules:

[96]

Face detection

This module is responsible to extract bounding box of face which includes coordinate values of top left and bottom right of facial area. This research implemented it using three online face detention algorithms, witch are put in series in order to increase the confident detection ratio. These three methods are respectively as follows:

- Skin detection followed by biggest contour (connected component)extractionREF
- HaarCascade object detection REF
- Classic Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid and sliding window detection scheme. (dlib face detector) REF

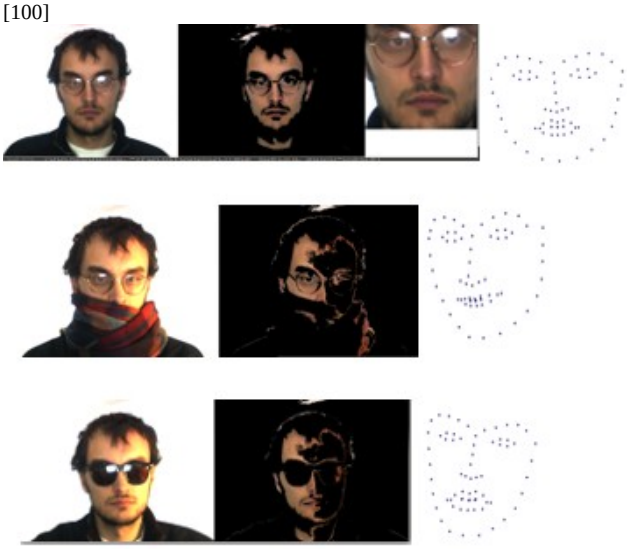
[97] Figure 1 illustrates from left to right a) normal face b)skin segmentation, and c) result of face detector.

Face Key point Extractor and Alignment

[98] After finding face bounding box in face detection module, the next step is finding the location of different facial features (e.g. corners of the eyes, eyebrows, the mouth,and the tip of the nose etc) accurately. Facial feature detection is also referred to “facial landmark detection”, “facial key point detection” and “face alignment” in the literature.REF In this research alignment modulemakes the eyes and nose appear in similar locations using a simple 2D affine transformation then scales frontal face image for input into a neural network. There are two based for making landmarks: “eyes and nose “and “eyes and bottom lips”. Land marks are shown in the right side of figure 2. Additionally, figure 2 proves even face is not detect by face detectors such as harrcascade due to have occlusion; face bounding box can be extractedusingdlib's landmark estimation.

[99] Figure 2 middle and bottom shows how face detection process has been affected by occlusion.

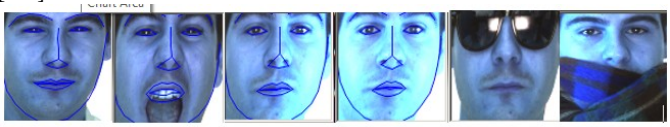
[100]



[101]

How ever **figure 3 illustrates that exaggerated occlusion prevents to landmark all facial features.**

[102]



[103]Generally, the following mathematical relations are satisfied between facial components to get a better result in detection the face affected by uncontrolled conditions assuming:

[104] $face\ aspect\ ratio\ (w/h)\ is\ around\ 0.125$

[105] $approximately\ ¿\ is\ located\ at\ (0.16\,w, 0.14\,h)$

[106] $face\ centre\ is\ centre\ of\ bounding\ box\ rectangle$

[107] $(x, y) = most\ top\ ¿\ coordinate\ value$

[108] $(w, h) = face\ dimension$

[109] $x_{face}\ center = x + w/2$

[110] $open\ both_{¿} = (0.7\,w, 0.3\,h)$

[111] $close\ both_{¿} = (0.6\,w, 0.25\,h)$

[112]

approximately closed eye is located at $(0.1w, 0.2h)$

[113] $eyc_{¿} = ¿)$

[114]

[115] In such a case, that none of cascading modules be able to detect both eyes, based on its obtained center face area will be searched half by half. Adjust the right and left-eyes rectangle one by one because the face border was re adjust the right-eye rectangle, since it starts on the rightside of the image.

[116] open both_eyes _dimension=(0.7w,0.3h)

[117] close both_eyes _dimension=(0.6w,0.25h)

[118] approximately closed eye is located at (0.1w ,0.2 h)

[119] eyc_glasses_dimension=(0.95w ,0.4 h)

[120]

[121] in the most cases one side has been affected more than other side by illumination In that case, applying histogram equalization on the whole face makes one half dark and one half bright , however applying histogram equalization side by side although makes them same on average but makes aw sharp edge in the middle of the face and the left half and right half would be suddenly different. Thus three obtained image by histogram equalization the whole will be blended together for a smooth brightness transition. After detecting face image is ready for rotation,scaling and translation to adjust eyes positions . Utilizing the "Bilateral Filter" will be reduced pixel noise by smoothing the image, but keeping the sharp edges in the face

Face Recognition

[122]

[123] Since through verification the system does not have any knowledge of type of occlusion thus making occlusion dictionary is not feasible. Therefore, this research supposes any input test sample is occluded and tries to extract occlusion mask from it. In the next step occlusion mask will be applied to all members of all classes in train dataset to generate synthetic occluded face. Finally, the distance between HOG features of input test sample and each synthetic occluded trained data is calculated, the shortest distance represents the class of input sample. The following describes how to occlusion mask is extracted from input image.

[124] In the HSV representation of color, Hue determines the color, Saturation determines color intensity and Value determines the image lightness.

[125] In order to find occlusion and illumination any captured frame for verification purpose RGB aligned detected face image first must be converted to HSV. Then to isolate the colors multiple masks have been applied to HSV image.The following describes how to occlusion mask is extracted from input image.

[126]

[127]

$T_{image_i} = Trained\ Front\ Facial\ Normal\ image\ for\ i = 1, ..$

[128] $m = number\ of\ exist\ classes \in trained\ dataset$

[129]

$R_{image} = RGB\ Received\ Sample\ Image\ for\ verification\ capture$

[130]

capture automatically by assesement service request

[131] $Function\ Mask\ Extractor(RGBimg\ ae)$

[132] $RGBimage(RGB)\ Mask(HSV)$

[133] $Mask(HSV)\ Mask(RGB)$

[134] $Mask(RGB)\ Mask(YCRB)$

[135]

$Mask(RGBimage) = Mask(HSV) + Mask(RGB) + Mask(YRCI$

[136] Return $Mask(RGBimage)$

[137] $Distanc\ e_{R1} = Obsolete\ Diff($
$Mask\ Extractor(R_{image}) - R_{image})$

[138] $Distanc\ e_{R2} = Obsolete\ Diff($
$Mask\ Extractor(T_{image_i}) - R_{image})$

[139]

$Affecte\ d_{Received} = Bitwise \vee (Distanc\ e_{R1}, Distanc\ e_{R2})$

[140] $Distanc\ e_{T1} = Obsolete\ Diff($
$Mask\ Extractor(R_{image}) - T_{image_i})$

[141]

$Distanc\ e_{T2} = Obsolete\ Diff\ (Mask\ Extractor(T_{image_i}) - T_{image_i})$

[142] $Affecte\ d_{Normal_i} = Bitwise \vee ¿)$

[143]

$Ho\ g_{Ti} = HistogramGradiant(HistogramEqualization($ *Affecte*

[144]

$Ho\ g_R = HistogramGradiant(HistogramEqualization($ *Affecte*

[145] for
$Min(EuclidianDistanc\ e_i)\ 1 < i < m\ i\ represents\ the\ class\ whic$

[146]

[147] $\forall\ i = 1, \ldots, n\ n = number\ of\ classes$

[148] $T_1, T_2, \ldots, T_n\ describes\ all\ normal\ train\ samples$

[149]

$m_1, m_2, \ldots, m_n\ describes\ mask\ of\ normal\ train\ samples$

[150] $d_{1T_i} = T_i - m_i$

[151] $d_{2T_i} = T_i - m_R$

[152] $d_{1R} = R - m_R$

[153] $d_{2T_iR} = T_i - m_R$

[154]

[155]

$$d_{1T_i} \vee d_{2T_i} = \sum_{n=0}^{[\log_2(x)]} 2^n \left[\left[\left(\lfloor \frac{d_{1T_i}}{2^n} \rfloor \bmod 2 \right) + \left(\lfloor \frac{d_{2T_i}}{2^n} \rfloor \bmod 2 \right) + \left(\lfloor \frac{d_{1T_i}}{2^n} \rfloor \right. \right. \right. \quad (1)$$

[156]

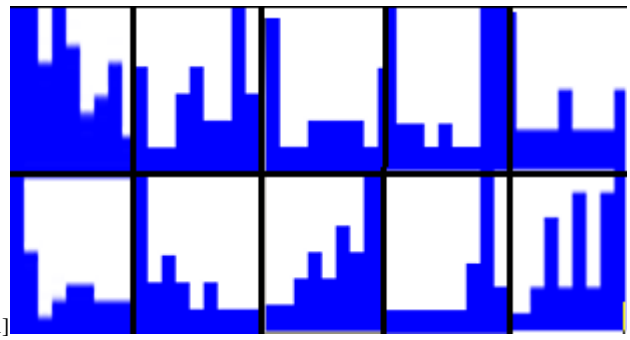
$$d_{1R} \vee d_{2T_i R} = \sum_{n=0}^{\lceil \log_2(x) \rceil} 2^n \left[\left(\left\lfloor \frac{d_{1R}}{2^n} \right\rfloor \bmod 2 \right) + \left(\left\lfloor \frac{d_{2T_i R}}{2^n} \right\rfloor \bmod 2 \right) \right.$$

[157](2)

[158]$\varepsilon_i =$ ¿histogram equalization (1) - histogram equalization (2) (3)

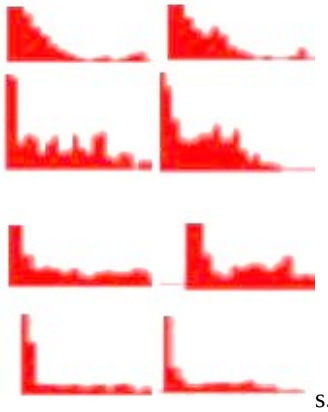
[159]The shortest distance (ε_i)indicates to the occluded training sample with same class as test sample.

[160]The figure number # visually illustrates formula (3).it respectively indicates $\varepsilon_i$of expression,illumination, sunglasses occlusion, scarf and scarf with illumination from left to right. When the top of figure is color histogram of original occlusion and the bottom is color histogram of synthetic occlusion.



[161] $\varepsilon_i \in$ *diffrent uncontroled condition figure* ¿

[162]Figure #+1 the demonstrates ε_i *value*in different condition supposing ε_i is distance of HOG feature vectors (1x64) rather than intensity value



s.

[163]

[164]

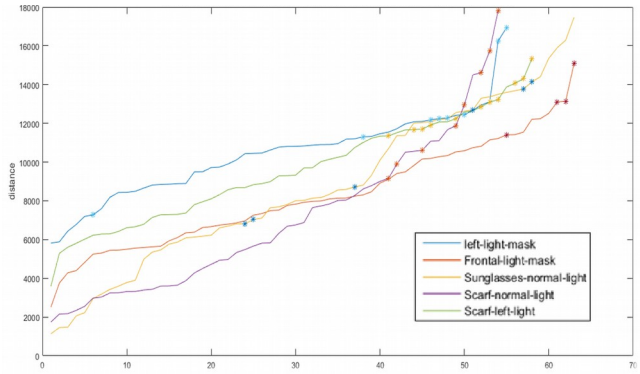
laplacian $\varepsilon_i \in$ *diffrent uncontroled condition figure*

¿

¿

[165]

[166]`



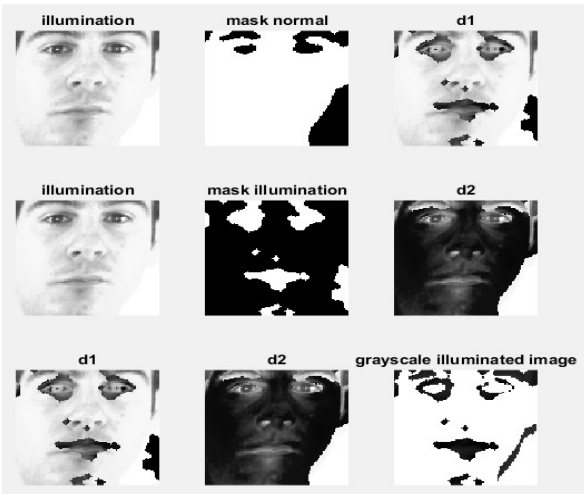
ε_i *range* $\in$ *diffrent uncontroled condition figure*

[167]

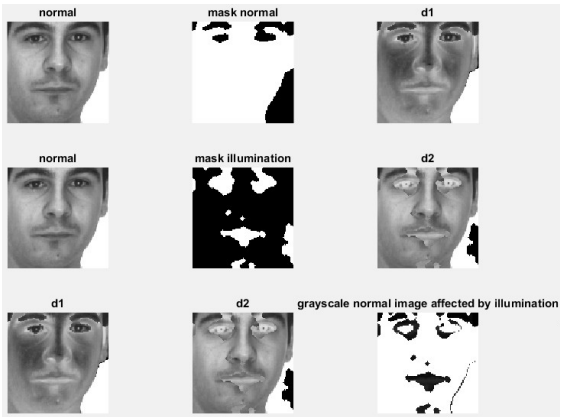
¿

¿

[168]Figure indicates that $\varepsilon_i$in failed condition ε_i has large value



[169]



EPERIMENTAL RESULTS AND EVALUATION

[170]this section demonstrates some experimental results of implementationproposed method. As stated (paint engineering) before in this study all uncontrolled conditions can be considered as occlusion. Therefore (paint engineering folan figure from maghale folan) is used to illustrate comparison the relation between recognition accuracy and the percentage occlusion in different conditions added some coordinate value to these figures.

[171]It can be concluded however it seems that illumination affects more than others but is accuracy is still high.

[172] By the result the worst case scenario is related to occluded face by scarf and affected by illumination at the same time.

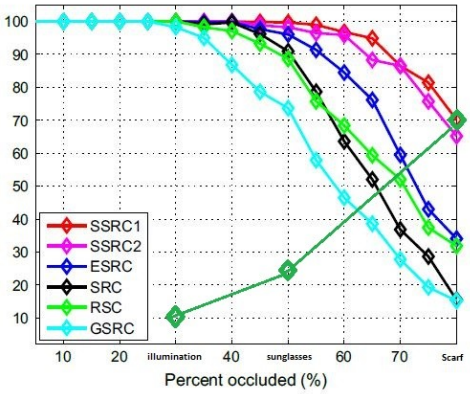
CONCOLUSION

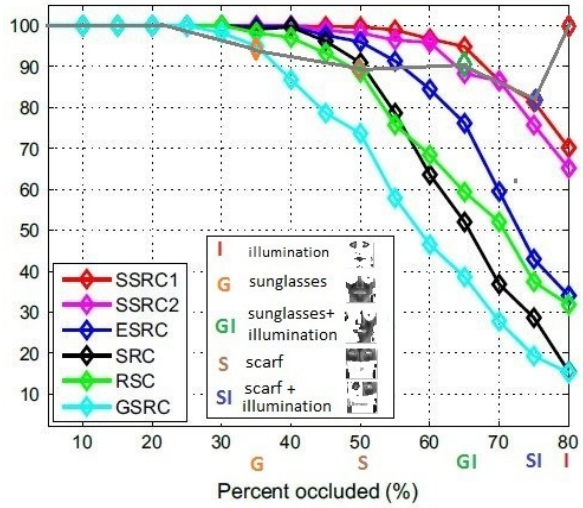
[173] the research undertaken is part of ongoing study to address the issues regarding occlusion in real time face recognition in small sample size dataset. The proposed method is based on finding occlusion mask, adding it to normal trained face in dataset and measuring the distance between synthetic and original occluded images. Further development will conduct to design a deep learning network for this purses.

[174]

[175]

[176]





[177]

[178]

[179]

[180]

[181]

[182]

[183]

[184]

[185]

[186]

[187]

[188]

[189]

[190]

[191]

[192]

[193]

- **SRC : Robust face recognition via sparse representation**, J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, IEEE Trans. Pattern Anal. **(2009)**
- **Extended SRC: undersampled face recognition via intraclass variant dictionary**, W. Deng, J. Hu, J. Guo, IEEE Trans. Pattern Anal. **(2012)**
- **SSRC: Robust face recognition via occlusion dictionary learning** Weihua Ou, Xinge You, Dacheng Tao, Pengyue Zhang, YuanyanTang, ZiqiZhu Elsevier Pattern Recognition journal **(2014)**
- **Structured occlusion coding for robust face recognition** Yandong Wen, Weiyang Liu, Meng Yang ,Neurocomputing 178 **(2016)**
- **A Survey on Face Detection in the wild: past, present and future** Computer Vision and Image Understanding. September **2015.**
- "**Face Recognition Methods & Applications**," B. M. DN Parmar, arxiv.org, **2014.**
- "**Human Detection from Images and Videos: A Survey“** D.Nguyen, W. Li and Philip O. Ogunbona, Pattern Recognition,Elsevier, **2015.**
- **GSRC: Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary** Meng

Yang , Lei Zhang , Simon C.K. Shiu ,Pattern Recognition **(2013)**

- **Deep learning in neural networks: An overview ,**J Schmidhuber , Neural networks, 2015 - Elsevier